Maths for ML

Nipun Batra and teaching staff

IIT Gandhinagar

August 17, 2025

Maths for ML

1. Given a vector of $\boldsymbol{\epsilon}$, we can calculate $\sum \epsilon_i^2$ using $\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}$

Maths for ML

1. Given a vector of $\boldsymbol{\epsilon}$, we can calculate $\sum \epsilon_i^2$ using $\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_N \end{bmatrix}_{N\times 1}$$

Maths for ML

1. Given a vector of $\boldsymbol{\epsilon}$, we can calculate $\sum \epsilon_i^2$ using $\boldsymbol{\epsilon}^T\boldsymbol{\epsilon}$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \ldots \\ \epsilon_N \end{bmatrix}_{N\times 1}$$

$$\boldsymbol{\epsilon}^T = \begin{bmatrix} \epsilon_1, \epsilon_2, \ldots, \epsilon_N \end{bmatrix}_{1\times N}$$

Maths for ML

1. Given a vector of $\boldsymbol{\epsilon}$, we can calculate $\sum \epsilon_i^2$ using $\boldsymbol{\epsilon}^T \boldsymbol{\epsilon}$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \ldots \\ \epsilon_N \end{bmatrix}_{N \times 1}$$

$$\boldsymbol{\epsilon}^T = \begin{bmatrix} \epsilon_1, \epsilon_2, \ldots, \epsilon_N \end{bmatrix}_{1 \times N}$$

$$\boldsymbol{\epsilon}^T \boldsymbol{\epsilon} = \sum \epsilon_i^2$$

Maths for ML

2.

$$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$$

Maths for ML

2. $$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$$

3. For a scalar s $$s = s^T$$

Maths for ML

4. Derivative of a scalar s wrt a vector $\boldsymbol{\theta}$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}$$

Maths for ML

4. Derivative of a scalar s wrt a vector $\boldsymbol{\theta}$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}$$

$$\frac{\partial s}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial s}{\partial \theta_1} \\ \frac{\partial s}{\partial \theta_2} \\ \vdots \\ \frac{\partial s}{\partial \theta_N} \end{bmatrix}$$

Linear Functions: Row Vector Times Column Vector

Definition: Setup

Configuration:

- $\boldsymbol{A}$ is a row vector $(1 \times n$ matrix$)$

Linear Functions: Row Vector Times Column Vector

Definition: Setup

Configuration:

- $\boldsymbol{A}$ is a row vector ($1 \times n$ matrix)
- $\boldsymbol{\theta}$ is a column vector ($n \times 1$ matrix)

Linear Functions: Row Vector Times Column Vector

Definition: Setup

Configuration:

- $\boldsymbol{A}$ is a row vector ($1 \times n$ matrix)
- $\boldsymbol{\theta}$ is a column vector ($n \times 1$ matrix)
- $\boldsymbol{A}^T\boldsymbol{\theta}$ produces a scalar

Linear Functions: Row Vector Times Column Vector

Definition: Setup

Configuration:

- $\boldsymbol{A}$ is a row vector ($1 \times n$ matrix)
- $\boldsymbol{\theta}$ is a column vector ($n \times 1$ matrix)
- $\boldsymbol{A}^T\boldsymbol{\theta}$ produces a scalar

Example: Concrete Example

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}, \quad \boldsymbol{A} = \begin{bmatrix} A_1 & A_2 \end{bmatrix}_{1\times 2}$$

Linear Functions: Row Vector Times Column Vector

Key Points

Matrix Multiplication Result

$$\boldsymbol{A\theta} = A_1\theta_1 + A_2\theta_2$$

Linear Functions: Row Vector Times Column Vector

Key Points

Matrix Multiplication Result

$$\boldsymbol{A\theta} = A_1\theta_1 + A_2\theta_2$$

This is a scalar! (Linear combination of parameters)

Linear Functions: Row Vector Times Column Vector

Key Points

Matrix Multiplication Result

$$\boldsymbol{A\theta} = A_1\theta_1 + A_2\theta_2$$

This is a scalar! (Linear combination of parameters)

Important: ML Relevance

This form appears everywhere in ML:

Linear Functions: Row Vector Times Column Vector

Key Points

Matrix Multiplication Result

$$\boldsymbol{A\theta} = A_1\theta_1 + A_2\theta_2$$

This is a scalar! (Linear combination of parameters)

Important: ML Relevance

This form appears everywhere in ML:

- Linear regression: $\boldsymbol{w}^T\boldsymbol{x}$
- Neural networks: $\boldsymbol{w}^T\boldsymbol{h} + b$
- Loss functions: $\boldsymbol{c}^T\boldsymbol{\theta}$

Gradient of Linear Function: Key Result

Key Points

Computing the Gradient

Gradient of Linear Function: Key Result

Key Points

Computing the Gradient

Goal: Find $\frac{\partial \boldsymbol{A\theta}}{\partial \boldsymbol{\theta}}$ where $\boldsymbol{A\theta} = A_1\theta_1 + A_2\theta_2$

Gradient of Linear Function: Key Result

Key Points

Computing the Gradient
Goal: Find $\frac{\partial \boldsymbol{A\theta}}{\partial \boldsymbol{\theta}}$ where $\boldsymbol{A\theta} = A_1\theta_1 + A_2\theta_2$

Example: Step-by-Step Calculation

$$\frac{\partial \boldsymbol{A\theta}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial}{\partial \theta_1}(A_1\theta_1 + A_2\theta_2) \\ \frac{\partial}{\partial \theta_2}(A_1\theta_1 + A_2\theta_2) \end{bmatrix}$$

Gradient of Linear Function: Key Result

Key Points

Computing the Gradient

Goal: Find $\frac{\partial \boldsymbol{A\theta}}{\partial \boldsymbol{\theta}}$ where $\boldsymbol{A\theta} = A_1\theta_1 + A_2\theta_2$

Example: Step-by-Step Calculation

$$\frac{\partial \boldsymbol{A\theta}}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial}{\partial \theta_1}(A_1\theta_1 + A_2\theta_2) \\ \frac{\partial}{\partial \theta_2}(A_1\theta_1 + A_2\theta_2) \end{bmatrix}$$

$$= \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}_{2\times 1} = \boldsymbol{A}^T$$

Gradient of Linear Function: Key Result

Important: Fundamental Rule

$$\boxed{\frac{\partial \boldsymbol{A\theta}}{\partial \boldsymbol{\theta}} = \boldsymbol{A}^T}$$

Gradient of Linear Function: Key Result

Important: Fundamental Rule

$$\boxed{\frac{\partial \boldsymbol{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \boldsymbol{A}^T}$$

This is one of the most important rules in ML optimization!

Gradient of Linear Function: Key Result

Important: Fundamental Rule

$$\boxed{\frac{\partial \boldsymbol{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \boldsymbol{A}^T}$$

This is one of the most important rules in ML optimization!

Definition: Intuition

Why $\boldsymbol{A}^T$?

Gradient of Linear Function: Key Result

Important: Fundamental Rule

$$\boxed{\frac{\partial \boldsymbol{A}\boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \boldsymbol{A}^T}$$

This is one of the most important rules in ML optimization!

Definition: Intuition

Why $\boldsymbol{A}^T$? Each component of the gradient equals the coefficient of the corresponding parameter in the linear function.

Quadratic Forms and Their Derivatives

Quadratic Forms: Introduction

Definition: Quadratic Form Derivative Rule

Key Result: For matrix $\boldsymbol{Z}$ of form $\boldsymbol{X}^T\boldsymbol{X}$:

Quadratic Forms: Introduction

Definition: Quadratic Form Derivative Rule

Key Result: For matrix $\boldsymbol{Z}$ of form $\boldsymbol{X}^T\boldsymbol{X}$:

$$\frac{\partial}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^T \boldsymbol{Z} \boldsymbol{\theta}) = 2\boldsymbol{Z}^T\boldsymbol{\theta}$$

Quadratic Forms: Introduction

Definition: Quadratic Form Derivative Rule

Key Result: For matrix $\boldsymbol{Z}$ of form $\boldsymbol{X}^T\boldsymbol{X}$:

$$\frac{\partial}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^T \boldsymbol{Z} \boldsymbol{\theta}) = 2\boldsymbol{Z}^T \boldsymbol{\theta}$$

Example: Understanding $\boldsymbol{X}^T\boldsymbol{X}$ Matrices

Quadratic Forms: Introduction

Definition: Quadratic Form Derivative Rule

Key Result: For matrix $\boldsymbol{Z}$ of form $\boldsymbol{X}^T\boldsymbol{X}$:

$$\frac{\partial}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}^T \boldsymbol{Z} \boldsymbol{\theta}) = 2\boldsymbol{Z}^T \boldsymbol{\theta}$$

Example: Understanding $X^T X$ Matrices

Starting with:

$$\boldsymbol{X} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \boldsymbol{X}^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

Quadratic Forms: Introduction

Key Points

Computing $\boldsymbol{Z} = \boldsymbol{X}^T \boldsymbol{X}$

Quadratic Forms: Introduction

Key Points

Computing $\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X}$

$$\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}_{2\times 2}$$

Quadratic Forms: Introduction

Key Points

Computing $\boldsymbol{Z} = \boldsymbol{X}^T \boldsymbol{X}$

$$\boldsymbol{Z} = \boldsymbol{X}^T \boldsymbol{X} = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}_{2 \times 2}$$

Important: Symmetric Property

Quadratic Forms: Introduction

Key Points

Computing $\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X}$

$$\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix}_{2\times 2}$$

Important: Symmetric Property

Key Observation: $Z_{ij} = Z_{ji} \Rightarrow \boldsymbol{Z}^T = \boldsymbol{Z}$ (symmetric matrix)

Maths for ML

Let

$$\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

Maths for ML

Let

$$\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

Maths for ML

Let

$$\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e\theta_1 + f\theta_2 \\ f\theta_1 + g\theta_2 \end{bmatrix}_{2\times 1}$$

Maths for ML

Let

$$\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e\theta_1 + f\theta_2 \\ f\theta_1 + g\theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2$$

Maths for ML

Let

$$\boldsymbol{Z} = \boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2}$$

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e & f \\ f & g \end{bmatrix}_{2\times 2} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}_{1\times 2} \begin{bmatrix} e\theta_1 + f\theta_2 \\ f\theta_1 + g\theta_2 \end{bmatrix}_{2\times 1}$$

$$\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta} = e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2$$

The term $\boldsymbol{\theta}^T\boldsymbol{Z}\boldsymbol{\theta}$ is a scalar.

Maths for ML

$$\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\theta}^T \boldsymbol{Z} \boldsymbol{\theta} = \frac{\partial}{\partial \theta} (e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2)$$

Maths for ML

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\theta}^T \boldsymbol{Z} \boldsymbol{\theta} &= \frac{\partial}{\partial \theta}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \\
&= \begin{bmatrix} \frac{\partial}{\partial \theta_1}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \\ \frac{\partial}{\partial \theta_2}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \end{bmatrix}
\end{aligned}
$$

Maths for ML

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\theta}^T \boldsymbol{Z} \boldsymbol{\theta} &= \frac{\partial}{\partial \theta}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \\
&= \begin{bmatrix} \frac{\partial}{\partial \theta_1}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \\ \frac{\partial}{\partial \theta_2}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \end{bmatrix} \\
&= \begin{bmatrix} 2e\theta_1 + 2f\theta_2 \\ 2f\theta_1 + 2g\theta_2 \end{bmatrix} = 2 \begin{bmatrix} e & f \\ f & g \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_1 \end{bmatrix}
\end{aligned}
$$

Maths for ML

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\theta}^T \boldsymbol{Z} \boldsymbol{\theta} &= \frac{\partial}{\partial \theta}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \\
&= \begin{bmatrix} \frac{\partial}{\partial \theta_1}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \\ \frac{\partial}{\partial \theta_2}(e\theta_1^2 + 2f\theta_1\theta_2 + g\theta_2^2) \end{bmatrix} \\
&= \begin{bmatrix} 2e\theta_1 + 2f\theta_2 \\ 2f\theta_1 + 2g\theta_2 \end{bmatrix} = 2 \begin{bmatrix} e & f \\ f & g \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_1 \end{bmatrix} \\
&= 2\boldsymbol{Z}\boldsymbol{\theta} = 2\boldsymbol{Z}^T\boldsymbol{\theta}
\end{aligned}
$$

Matrix Rank and Invertibility

Matrix Rank: Fundamental Concept

Definition: What is Matrix Rank?

Matrix Rank: Fundamental Concept

Definition: What is Matrix Rank?

Rank = Maximum number of linearly independent rows (or columns)

Matrix Rank: Fundamental Concept

Definition: What is Matrix Rank?

Rank = Maximum number of linearly independent rows (or columns)

Key Points

Two Equivalent Perspectives For an $r \times c$ matrix:

Matrix Rank: Fundamental Concept

Definition: What is Matrix Rank?

Rank = Maximum number of linearly independent rows (or columns)

Key Points

Two Equivalent Perspectives For an $r \times c$ matrix:

- **Row perspective:** r row vectors, each with c elements

Matrix Rank: Fundamental Concept

Definition: What is Matrix Rank?

Rank = Maximum number of linearly independent rows (or columns)

Key Points

Two Equivalent Perspectives For an $r \times c$ matrix:

- **Row perspective:** r row vectors, each with c elements
- **Column perspective:** c column vectors, each with r elements

Matrix Rank: Fundamental Concept

Example: Maximum Rank Rules

- If $r < c$: Maximum rank $= r$ (more columns than rows)

Matrix Rank: Fundamental Concept

Example: Maximum Rank Rules

- If $r < c$: Maximum rank $= r$ (more columns than rows)
- If $r > c$: Maximum rank $= c$ (more rows than columns)

Matrix Rank: Fundamental Concept

Example: Maximum Rank Rules

- If $r < c$: Maximum rank $= r$ (more columns than rows)
- If $r > c$: Maximum rank $= c$ (more rows than columns)
- If $r = c$: Maximum rank $= r = c$ (square matrix)

Matrix Rank: Fundamental Concept

Example: Maximum Rank Rules

- If $r < c$: Maximum rank $= r$ (more columns than rows)
- If $r > c$: Maximum rank $= c$ (more rows than columns)
- If $r = c$: Maximum rank $= r = c$ (square matrix)

Important: ML Relevance

Why rank matters:

- Determines if matrix is invertible

Matrix Rank: Fundamental Concept

Example: Maximum Rank Rules

- If $r < c$: Maximum rank $= r$ (more columns than rows)
- If $r > c$: Maximum rank $= c$ (more rows than columns)
- If $r = c$: Maximum rank $= r = c$ (square matrix)

Important: ML Relevance

Why rank matters:

- Determines if matrix is invertible
- Affects uniqueness of solutions

Matrix Rank: Fundamental Concept

Example: Maximum Rank Rules

- If $r < c$: Maximum rank $= r$ (more columns than rows)
- If $r > c$: Maximum rank $= c$ (more rows than columns)
- If $r = c$: Maximum rank $= r = c$ (square matrix)

Important: ML Relevance

Why rank matters:

- Determines if matrix is invertible
- Affects uniqueness of solutions
- Critical for understanding overfitting

Maths for ML: Matrix Rank

- Given a matrix $\boldsymbol{A}$:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

Maths for ML: Matrix Rank

- Given a matrix $\boldsymbol{A}$:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

- What is the rank?

Maths for ML: Matrix Rank

- Given a matrix $\boldsymbol{A}$:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

- What is the rank?
- $r = c = 3$. Thus, rank is $<= 3$

Maths for ML: Matrix Rank

- Given a matrix $\boldsymbol{A}$:

$$\begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 1 \\ 2 & 7 & 8 \end{bmatrix}$$

- What is the rank?
- $r = c = 3$. Thus, rank is $<= 3$
- Row 3 can be written as: 3 times Row 1 + 2 times Row 1. Thus, Row 3 is linearly dependent on Row 1 and 2. Thus, rank($\boldsymbol{A}$)=2

Maths for ML: Matrix Rank

What is the rank of

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & 4 & 4 \\ 3 & 4 & 8 & 0 \end{bmatrix}$$

Maths for ML: Matrix Rank

What is the rank of

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & 4 & 4 \\ 3 & 4 & 8 & 0 \end{bmatrix}$$

Since $\boldsymbol{X}$ has fewer rows than columns, its maximum rank is equal to the maximum number of linearly independent rows. And because neither row is linearly dependent on the other row, the matrix has 2 linearly independent rows; so its rank is 2.

Maths for ML: Matrix Inverse

Suppose $\boldsymbol{A}$ is an $n \times n$ matrix. The inverse of $\boldsymbol{A}$ is another $n \times n$ matrix, denoted $\boldsymbol{A}^{-1}$, that satisfies the following conditions.

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

where $\boldsymbol{I}_n$ is the identity matrix.

Maths for ML: Matrix Inverse

Suppose $\boldsymbol{A}$ is an $n \times n$ matrix. The inverse of $\boldsymbol{A}$ is another $n \times n$ matrix, denoted $\boldsymbol{A}^{-1}$, that satisfies the following conditions.

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

where $\boldsymbol{I}_n$ is the identity matrix.

Below, with an example, we illustrate the relationship between a matrix and its inverse.

Maths for ML: Matrix Inverse

Suppose $\boldsymbol{A}$ is an $n \times n$ matrix. The inverse of $\boldsymbol{A}$ is another $n \times n$ matrix, denoted $\boldsymbol{A}^{-1}$, that satisfies the following conditions.

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

where $\boldsymbol{I}_n$ is the identity matrix.
Below, with an example, we illustrate the relationship between a matrix and its inverse.

$$\begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0.8 & -0.2 \\ -0.6 & 0.4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0.8 & -0.2 \\ -0.6 & 0.2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

A square matrix that has an inverse is said to be nonsingular or invertible; a square matrix that does not have an inverse is said to be singular.

Maths for ML: Matrix Inverse

There are two ways to determine whether the inverse of a square matrix exists.

- If the rank of an $n \times n$ matrix is less than n, the matrix does not have an inverse.
- When the determinant for a square matrix is equal to zero, the inverse for that matrix does not exist.

A square matrix that has an inverse is said to be nonsingular or invertible; a square matrix that does not have an inverse is said to be singular.
Not every square matrix has an inverse; but if a matrix does have an inverse, it is unique.